ABSTRACT

Described are techniques for processing data sets produced by analyzing a sample. The input data set is represented as rows of intensities over time for a particular mass to charge (m/z) range. A correlation matrix is produced in which each row of the input data set is correlated with every other row in the input data set. The correlation matrix is clustered or grouped such that those highly correlated m/z ranges are included in the same group. A set of one or more scans is selected for each group representing periods of interest within each group. Using the m/z values included in each cluster, a resultant sample spectra is created for each of the selected scans. The processing techniques may be used to identify parent and related fragment ions in the input data set and as a preprocessor producing resultant sampled spectra used as input to subsequent processing.